

EXPLOITING AN EVENT BASED STATE ESTIMATOR IN PRESENCE OF SPARSE MEASUREMENTS IN VIDEO ANALYTICS

Pietro Morerio, Mattia Pompei, Lucio Marcenaro, Carlo S. Regazzoni

DITEN - University of Genova
Via Opera Pia 11A 16145 Genoa - Italy
pmorerio@ginevra.dibe.unige.it

ABSTRACT

Recently, a Bayesian estimator with a hybrid update was developed [1], based on a mathematical formulation of sampling. Such an Event Based State Estimator (EBSE) allows for a stable synchronous state estimate, relying on asynchronous measurements. Usefulness of such a filter comes with its approximate analytic formulation, which is attainable given a send-on-delta sampling strategy. We argue that such a formulation can be extended to cope with a failing detector in case the filter is used for tracking. The basic idea is to approach the issue as a package loss problem, where a missed target is assimilated to a lost package. More in detail, we propose that this approach can be exploited in video tracking, where faulty detectors are commonplace. We show how tracking performance with a poor pedestrian detector, failing to recognize its target, can improve with respect to standard Kalman filter.

Index Terms— Bayesian filtering, EBSE, Hybrid update, Package loss, Sparse measurements, Failing detector

1. INTRODUCTION

A stable event-based Bayesian state estimator was developed in [1], based on a mathematical formulation of sampling. Such an estimator implements a hybrid update, which allows for a synchronous estimate, while relying on asynchronous measurement events. Stability is obtained by updating the estimation results not only when a new measurement (event) is available, but also synchronously in time, i.e. when no measurement is received. At synchronous time instants, the update is based on the implicit information still available even when not receiving a new measurement. The simple, yet effective idea of updating the EBSE at its synchronous time instants, by exploiting a bounded measurement set, turns out to be sufficient for the derivation of an asymptotic bound on its error-covariance, regardless of the sampling strategy employed for defining events.

Most currently available analyses are limited to synchronous estimators, such as the Kalman filter (KF) [2] and its non-linear and non-Gaussian variations. An extension to package loss is provided by Sinopoli et al. in [3]. Here the (discrete) arrival of the observations is modelled as a random process whose parameters are related to the characteristics of an unreliable communication channel. The statistical convergence of the expected estimation error covariance of the filter is proved to be dependent on the probability of arrival of the package λ , given a system dynamic. The threshold λ_c , below which the expectation of the estimation error covariance is unbounded, is worked out in closed form. The authors also claim that this result can be interpreted as a manifestation of the *uncertainty threshold principle*.

Later on, Suh et al. proposed in 2006 a state estimation algorithm [4] which incorporates a send-on-delta strategy [5] (Figure 1(a)) in the discrete KF, to reduce data traffic in networks. Such a strategy copes with a sensor whose data are transmitted only if its output changes more than some value Δ . At synchronous time instants, no acquisition of sensor data implies that such a threshold is not exceeded. Yet, this information allows to increase estimation performance by reducing estimation errors. The method was later improved in [6]. Noticeably, the results developed in [1] trivially apply to this method, which turns out to be a particular case of the EBSE with a send-on-delta event sampling strategy and a rougher approximation in the formulation of the likelihood, as pointed out in [7].

We argue that the issue of *state estimation relying on a failing detector* is equivalent to the modelling of a virtual channel with package loss, and can be therefore approached by means of an EBSE-based filter. Here the event is represented by the detection, as it takes place. However, the EBSE has a very general formulation, which on the one hand allows to work out general results (such as its stability), which hold regardless of the sampling strategy adopted; on the other hand one is forced to choose a sampling method in order to get a practical tool to be applied in real applications. Unfortunately, not every sampling strategy allows for a straightforward analytical formulation of the EBSE. For instance, Integral Sampling (Figure 1(b)) does not allow a computation of the bounded measurement set requested for the likelihood formulation. In order to obtain an applicable tool, we rely on the analytical formulation obtained in [7] for the send-on-delta EBSE. We show how this can be applied to the problem of video tracking with failing detectors by making the correct assumptions. In particular we show results for a HOG-descriptor-based pedestrian detector [8], which can fail to recognize human targets in heterogeneous sets of frames. In particular, we show how tracking performance with a poor pedestrian detector, failing to identify its target, can improve with respect to the standard KF, even by relaxing some of the assumption done for the EBSE. In this sense, we deal with *sparse* observations.

The rest of this work is organized as follows: section 2 briefly explains the EBSE filter and its extension to a failing detector framework with a send-on-delta sampling strategy. Section 3 describes the experimental setup and discuss results. Conclusions and final remarks are given in section 4.

2. EBSE

A comprehensive mathematical formulation of the EBSE is given in [7], together with the demonstration of its stability. We here only state the problem and briefly sketch those parts of the method which

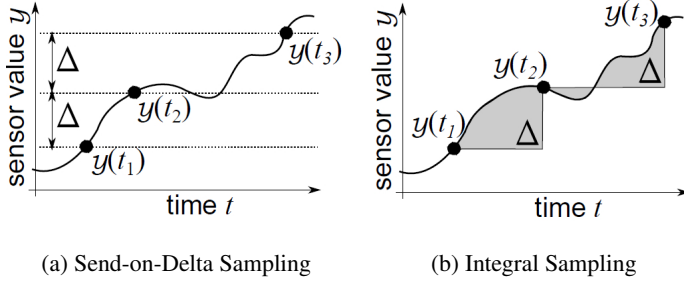


Fig. 1. The k -th event instant is triggered by the condition $|y(t) - y(t_{k-1})| \geq \Delta$ for the send-on-delta sampling strategy and by the condition $\int_{t_{k-1}}^t |y(\tau) - y(t_{k-1})| d\tau \geq \Delta$ for integral sampling [7].

are essential for the understanding of what will be presented in the following.

Event based state estimation deals with a sensor node exchanging measurements with a centralized state estimator. Measurements are exchanged only at instants which correspond to some *events* (i.e. asynchronously). However, estimations are often required synchronously in time (i.e. at each t_k).

Consider the time-discrete process

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad (1)$$

$$y_k = Cx_k + Du_k + v_k, \quad (2)$$

where the random variables w_{k-1} and v_k are the process and measurement noise respectively. The classical Bayesian estimator aims at computing the density $p(x_k|Y_k)$ by first *predicting* the *a priori* pdf

$$p(x_k|Y_{k-1}) = \int_{-\infty}^{+\infty} p(x_k|x_{k-1})p(x_{k-1}|Y_{k-1})dx_{k-1}, \quad (3)$$

where $Y_k = \{y_0, y_1, \dots, y_k\}$,

by exploiting the knowledge on the system process $p(x_k|x_{k-1})$ (inherent in eq. 1) and on the last estimation. The estimation is then updated with the new measure at time step k

$$p(x_k|Y_k) = \frac{p(y_k|x_k)p(x_k|Y_{k-1})}{p(y_k|Y_{k-1})}. \quad (4)$$

The problem has very well known solutions, such as the already mentioned KF (in the case of linearity of equations 1 and 2 and Gaussian noise) and its many extensions for non-linear equations and non-Gaussian noise (Extended KF, Unscented KF, Cubature Filter [9], Particle Filter [10]).

The solution to the problem of how to synchronously perform the update (eq. 4) when the sensor node does not send any measure relies on what can be interpreted as an extension of the concept of "measure". One can think of a measure as of new information coming into the system, in order to compute eq. 4. Such incoming knowledge can either be the actual measurement $y(t)$ or some less informative data such as the bounded set $\mathcal{H}[e|t]$ of all allowable values that $y(t)$ may take at time t (after the last measure $y(t_{e-1})$): this will be explained more in details in section 2.1):

$$\mathcal{Y}_k := \begin{cases} \mathcal{H}[e|t] & \text{No measurement;} \\ y(t_e) & \text{Measurement event.} \end{cases} \quad (5)$$

From here, a unified update equation can be worked out based on the expression of the likelihood

$$p(\mathcal{Y}_k|x_k) = \int p(y_k|x_k)p(y_k \in \mathcal{Y}_k)dy_k \quad (6)$$

where it is reasonable to conjecture that

$$p(y_k \in \mathcal{Y}_k) := \begin{cases} \Pi_{\mathcal{H}[e|t_k]} & \text{No measurement;} \\ \delta(y_k - y(t_e)) & \text{Measurement event;} \end{cases} \quad (7)$$

where $\Pi_{\mathcal{H}}$ is the uniform distribution over \mathcal{H} and $\delta(y)$ is the Dirac delta.

From now on, we will give a slightly different formulation of the EBSE, based on switching variables. A comprehensive theoretical dissertation on switching Bayesian models is given in [11], together with an example. By introducing the binary variable $\alpha = 0, 1$, eq. 6 can be rewritten in a clearer way as

$$p(\mathcal{Y}_k|x_k, \alpha) = \int p(y_k|x_k) [\alpha \delta(y_k - y(t_e)) + (1 - \alpha) \Pi_{\mathcal{H}[e|t_k]}] dy_k, \quad (8)$$

where $\alpha = 0$ if no new measurement comes from the sensor node and $\alpha = 1$ if, on the opposite, a new measure is triggered. The switching variable α is known at each synchronous instant t_k , as we of course are aware of whether the detection takes place or not.

In order to have a consistent formulation which holds regardless of the event sampling strategy, $\Pi_{\mathcal{H}}$ is approximated with a summation of an arbitrary number of equally spaced Gaussians with the same covariance [12]. After the state pdf $p(x_k|\mathcal{Y}_k[0, k])$ is eventually computed, its single-Gaussian approximation is also worked out, in order to attain computational tractability and make it possible to investigate the asymptotic behaviour of the filter.

2.1. Sampling

Triggering a sample (event) can be based either on time $t \in \mathbb{R}$ or on measurement values $y(t) \in \mathbb{R}^n$. Indeed, the formulation of sampling is grounded on the set $\mathcal{H}[e] \subseteq \mathbb{R}^{n+1}$ of all allowable values that $(y(t), t)$ may take after the last measurement $y(t_{e-1})$ and before the next $y(t_e)$. The event instant t_e is generated by

$$t_e := \inf \left\{ t \in \mathbb{R}^+ \mid t > t_{e-1} \wedge \begin{pmatrix} y(t) \\ t \end{pmatrix} \notin \mathcal{H}[e] \right\} \quad (9)$$

With this definition, the previously introduced $\mathcal{H}[e|t]$ can be obtained by a more rigorous description

$$\mathcal{H}[e|t] := \left\{ y \in \mathbb{R}^n \mid \begin{pmatrix} y \\ t \end{pmatrix} \in \mathcal{H}[e] \right\} \quad (10)$$

One of the key points of the EBSE algorithm lies in eq. 7 and consists in working out the (approximated) uniform distribution over $\mathcal{H}[e|t]$. It is therefore essential to construct such a set for an operative implementation of the filter. Choosing a sampling strategy essentially defines $\mathcal{H}[e]$ and thus $\mathcal{H}[e|t] \forall t \in (t_{e-1}, t_e)$. For example, for a send-on-delta case (Figure 1(a)), $\mathcal{H}[e|t]$ is nothing but the interval $(y(t_{e-1}) - \Delta, y(t_{e-1}) + \Delta)$ for each t in the interval (t_{e-1}, t_e) , as it can also be easily guessed by the plot. On the other side, it is easy to show that Integral Sampling (Figure 1(b)) does not permit a computation of the bounded measurement set $\mathcal{H}[e|t]$ requested for the likelihood formulation, which turns out to be dependent on the (generally unknown) analytic expression of $y(t)$ in the integration domain (even assuming y is monotonic in such an interval, only an upper bound can be worked out).

2.2. Extension to failing detectors

As already pointed out, we want to apply the EBSE to deal with the issue of a failing detector, in order to achieve stable tracking in video analysis. To this end, we consider the detection algorithm itself as a sensor node. Then, instead of considering a failing detector, we can think of a perfect detector (sensor node) communicating with an estimator, but experiencing package loss. Package loss rate depends on the performance of the chosen detector in the specific application domain.

We mentioned before that a sampling strategy must be selected in order to explicitly find an implementation of the filter. However, this is not completely correct: a sampling strategy is *sufficient* to give a (not always analytical) characterization of the set $\mathcal{H}[e|t]$ at each time t , but not necessary. We have already pointed out the importance of specifying such a set at each time t_{e-1} in order to approximate the uniform distribution in eq. 8. However, we argue that one can always think of specifying $\mathcal{H}[e|t]$ "by hands", without necessarily deriving such specification from a given sampling strategy.

Namely, we refine in some sense the EBSE+send-on-delta algorithm for application to failing detectors by relaxing the assertion that the hybrid update is supported by a sampling strategy and simply rely on the assumption that at synchronous time instants, when no measurement arrives from the sensor node, its value lies within a bounded set nevertheless. The algorithm is still guaranteed to converge to a bounded error-covariance as can be straightforwardly derived from the asymptotic analysis of the EBSE proposed in [7]. It is there in fact proved that there exist an asymptotic bound on the largest (positive) eigenvalue of the error-covariance, provided that $\mathcal{H}[e|t]$ is bounded at each time t , even in the situation that no new measurement is received.

We thus have to provide $\mathcal{H}[e|t]$ to the EBSE algorithm at each time t . For a video tracking state estimation problem, it is common understanding (and reasonable to conjecture) that, given the position $y(t_k)$ of the tracked target at time instant t_k , its position can be located, at time t_{k+1} , within a $(y(t_k) - \Delta, y(t_k) + \Delta)$ interval for some *finite* (and appropriate) Δ . Such a span can be interpreted as the $(y(t_{e-1}) - \Delta, y(t_{e-1}) + \Delta)$ bounded set previously introduced for the send on delta strategy. However, as $\mathcal{H}[e|t]$ is not here derived from a sampling strategy, Δ is not an intrinsic parameter related to the sampling but must be estimated. Such an estimation must be heuristic, at least to some extent, depending mainly on the dynamics of the problem and on the geometry of the camera setup.

Looking at the problem from the reverse perspective, we can mathematically interpret $\mathcal{H}[e|t]$ as the set where there is an approximately uniform non-zero probability of having a measure at time t (cfr. eq. 7). A reasonable guess for such a set can be constructed from the system model equations (eq. 1, where we consider white Gaussian noise, i.e. $p(w) = N(0, Q)$), by calculating the difference

$$|x_k - x_{k-1}| = |(A - I)x_{k-1} + Bu_{k-1} + w_{k-1}|, \quad (11)$$

where the time $k - 1$ is here to be considered as the last time a measurement arrived, i.e. t_{e-1} . By taking the largest eigenvalues λ_A , λ_B and λ_Q of the matrices A , B and Q respectively, we can at least have an idea of the order of magnitude of a reasonable Δ .

$$\Delta \approx |(\lambda_A - 1)\hat{x}_{k-1}| + |\lambda_B u_{k-1}| + \lambda_Q, \quad (12)$$

We empirically choose Δ as twice this value, to be sure that the integral in eq. 8 is always non-zero. As already pointed out, $\mathcal{H}[e|t]$ is not here derived from a sampling strategy. Therefore, the next measurement event instant t_e is not triggered as in eq. 9 and there might be an overlapping of two sets at subsequent time instants as $\mathcal{H}[e|t]$

is overestimated. This can result in a less accurate estimation, yet the EBSE is guaranteed to converge, as the proof on its asymptotic behaviour relies only on the fact that $\mathcal{H}[e]$ is bounded.

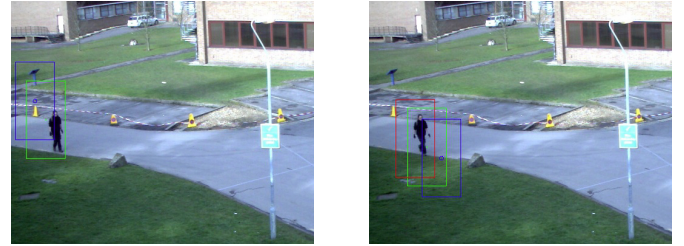
We have so far considered a situation where a single detection is missed and $\mathcal{H}[e|t]$ is to be computed once. However, a detector failing in a set of m consecutive frames is a common situation. In this case, $\mathcal{H}[e|t]$ has to be evaluated at each time t_k as long as a new measure arrives. The inherent knowledge of the number m of missed detections from the last measure can be here exploited to this end.

$$\mathcal{H}[e|t_k] = (y(t_{e-1}) - m \cdot \Delta, (y(t_{e-1}) + m \cdot \Delta) \quad (13)$$

where t_k is the time at which the m -th consecutive missed detection takes place.

3. RESULTS

We tested the extended EBSE on the benchmark video sequences from the 2009 PETS¹ dataset. A single camera view was tested, namely *View 1*; frame-rate of the sequence is $\sim 7[fps]$. Obviously, detection results do change with a different camera view and improve as the framing gets more and more lateral, because of the better defined pedestrian shape. However the point here is not about detection performance; in fact we welcome missed detections as a way of testing the estimation algorithm. Data association could be quite easily included in the presented algorithm, but at this stage the filter supports one target only. We thus had to select those clips of the sequence where a single person crosses the scene. Sample frames are shown in figure 2, together with the bounding boxes for detections, prediction and updates.



(a) Detection event

(b) Missed detection

Fig. 2. Sample frames and the EBSE at work, with bounding boxes: the *blue* rectangle depicts the predicted position; the *red* one shows the last detection (measure); the *green* box is the updated estimation.

For what concerns the testing detector, we employed one of the most common state of the art descriptors, namely Histograms of Oriented Gradients (HOG) features. Such cues were widely used for pedestrian detection in the last few years [14] and led to detector based on cascades of binary classifiers [15]. We exploited the C++ implementation of the open source HOG-based people detector supplied by the OpenCV computer libraries [16]. This gives a bounding box framing the pedestrian as an output. The state x_k here considered is the centre (x, y) of the detected bounding box, as outputted by the detector.

The algorithm was compared with a standard KF, which at each time step t_k estimates the state by first *predicting* its *a priori* pdf

¹Performance Evaluation of Tracking and Surveillance workshop [13].

(minus superscript) and then *updating* the mean and the covariance of its *a posteriori* pdf:

$$\begin{aligned}\hat{x}_k^- &= A\hat{x}_{k-1} + Bu_{k-1} & (a) \\ P_k^- &= AP_{k-1}A^T + Q & (b)\end{aligned}\quad (14)$$

$$\begin{aligned}\hat{x}_k &= \hat{x}_k^- + K_k(y_k - C\hat{x}_k^-) & (a) \\ P_k &= (I - K_kC)P_k^- & (b) \\ K_k &= P_k^-C^T(CP_k^-C^T + V)^{-1} & (c)\end{aligned}\quad (15)$$

where K_k is the Kalman gain, Q and R are the covariance matrices of $p(w) = N(0, Q)$ and $p(w) = N(0, R)$ respectively.

Such a filter (together with all its previously mentioned non-linear and non-Gaussian extensions) copes with a missing detection by utterly skipping the *measurement update* step (eq. 15). Basically, estimation is performed based on prediction only, as in dead-reckoning, and it is then subject to cumulative errors, as it can be easily guessed from the prediction equation for the state covariance, eq. 14(b). Skipping the update step does not allow P_k to diminish thanks to the Kalman gain, as can be deduced from eq. 15(b).

The most significant quantity to be studied is therefore the (a posteriori) covariance matrix P_k , which incidentally represents the second moment of the state distribution

$$P_k = E \left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T \right], \quad (16)$$

measuring the square discrepancy between the state and its expected value. Actually, we compare the traces $T(P_k)$ of the matrices, which, being exactly the sum of the eigenvalues, can give an idea of the maximum amount of dispersion of the distribution. Results for the most significant portions of the considered video sequences are given in Figures 3 and 4.

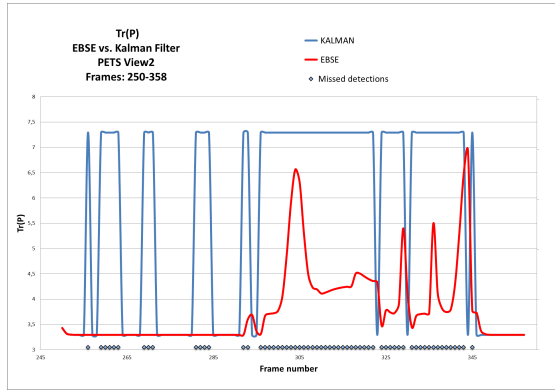


Fig. 3. $T(P_k)$ for KF and EBSE versus time.

The parameters set in equations 14 and 15 are as follows

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 1 & 0.35 \\ 0 & 1 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (17)$$

$$Q = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, R = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \quad (18)$$

As expected, the values of $T(P_k)$ are comparable, indeed almost identical, when detections take place. In fact, the two filters have the very same behaviour in case a measure arrives from the (virtual) sensor node. The two graphs had actually to be cut on the y axis for

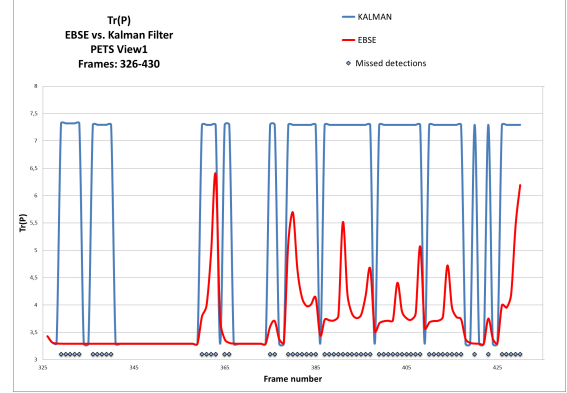


Fig. 4. $T(P_k)$ for KF and EBSE versus time.

readability. The higher the number of subsequent missed detection, the higher the peaks. It can be easily seen that, on the contrary, the value of $T(P_k)$ for the EBSE algorithm always remains bounded, supporting our initial guess that the EBSE could be heuristically extended beyond the sampling strategy approach.

As a final remark, we would like to point out that a discussion on false detections was not included here, as this phenomenon affects in the same way both the KF and the EBSE.

4. CONCLUSION

In this work we have proposed a possible extension of Event Based State Estimation for failing detectors. This is accomplished by considering a node framework where the link between the sensor node and the estimator experiences package loss. We suggest that the assumption that measures are triggered by a sampling strategy can be relaxed and the hybrid update of the estimator can simply rely on the assumption that at synchronous time instants, when no measurements arrive from the sensor node, there exists a bounded set in which there is uniform probability of finding the lost measure. Such a bounded set is upper bounded based on the dynamics of the problem. Namely, Even without considering a rigorous mathematical definition of sampling, but constructing a guess for the set H by means of eq. 13, the system attains stability.

Results show that the application to a failing HOG-based pedestrian detector in a video analysis framework is justified. The trace of the error covariance matrix remains bounded, meaning that asymptotic stability of the filter is accomplished. In this sense the proposed approach outperforms the standard KF-like estimation, which is subject to cumulative errors at synchronous time instants if no measures are sent to the estimator.

Future developments of this work will include an extension to the Joint Probability Data Association Filter for multiple target tracking in video analysis. Also, we are confident that a finer characterization of the set \mathcal{H} might be possible. Such a representation may recall the one described for the integral sampling, which is worth giving a further investigation. In fact, its mechanism, which includes time, is particularly suitable for a video surveillance state estimation problem, where duration of stay in certain areas becomes a relevant cue.

5. REFERENCES

- [1] J. Sijs and M. Lazar, "Event based state estimation with time synchronous updates," *Automatic Control, IEEE Transactions on*, vol. 57, no. 10, pp. 2650–2655, 2012.
- [2] R.E. Kalman, "A new approach to linear filtering and prediction problems," *T-ASME*, vol. 1960, pp. 35–45, March 1960.
- [3] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M.I. Jordan, and S.S. Sastry, "Kalman filtering with intermittent observations," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1453–1464, 2004.
- [4] Young-Soo Suh, Vinh Hao Nguyen, and Young-Shick Ro, "Networked estimation using a send-on-delta method," in *IEEE Industrial Electronics, IECON 2006 - 32nd Annual Conference on*, 2006, pp. 4622–4626.
- [5] Marek Miskowicz, "Send-on-delta concept: an event-based data reporting strategy," *Sensors*, vol. 6, pp. 49–63, 2006.
- [6] Vinh Hao Nguyen and Young Soo Suh, "Improving estimation performance in networked control systems applying the send-on-delta transmission method," *Sensors*, vol. 7, no. 10, pp. 2128–2138, 2007.
- [7] J. Sijs, *State estimation in networked systems*, Ph.D. thesis, Eindhoven Univ. Technol., Eindhoven, The Netherlands, 2012.
- [8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886–893 vol. 1.
- [9] I. Arasaratnam and Simon Haykin, "Cubature kalman filters," *Automatic Control, IEEE Transactions on*, vol. 54, no. 6, pp. 1254–1269, 2009.
- [10] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House Radar Library. Artech House, 2004.
- [11] Simone Chiappino, Lucio Marcenaro, Pietro Morerio, and Carlo Regazzoni, "Event based switched dynamic bayesian networks for autonomous cognitive crowd monitoring," in *Augmented Vision and Reality*, Augmented Vision and Reality, pp. 1–30. Springer Berlin Heidelberg, 2013.
- [12] H.W. Sorenson and D.L. Alspach, "Recursive bayesian estimation using gaussian sums," vol. 7, pp. 465–479, 1971.
- [13] James Ferryman and James L. Crowley, Eds., *Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS 2009*, 2009.
- [14] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886–893 vol. 1.
- [15] Hui-Xing Jia and Yu-Jin Zhang, "Fast human detection by boosting histograms of oriented gradients," in *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on*, 2007, pp. 683–688.
- [16] Open Source Computer Vision Library, "OpenCV," <http://opencv.org>.